AD_____

Award Number: W81XWH-04-1-0082

TITLE: PARK2, a Large common Fragile Site Gene, Is Part of a Stress Response Network in Normal Cells that is Disrupted During the Development of Ovarian Cancer

PRINCIPAL INVESTIGATOR: David I Smith, Ph.D.
Yu Zhu, M.D., Ph.D.

CONTRACTING ORGANIZATION: Mayo Clinic
Rochester MN 55905

REPORT DATE: January 2008

TYPE OF REPORT: Final Addendum

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-01-2008	Final Addendum	15 Dec 06 – 14 Dec 07

4. TITLE AND SUBTITLE
PARK2, a Large common Fragile Site Gene, Is Part of a Stress Response Network in Normal Cells that is Disrupted During the Development of Ovarian Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0082

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David I Smith, Ph.D.; Yu Zhu, M.D., Ph.D.

E-Mail: smith.david@mayo.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Mayo Clinic
Rochester MN 55905

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Please see Abstract on page 4.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	18	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

ABSTRACT

PARK2 (Parkin) is a common fragile site (CFS) gene. We examined Parkin in primary ovarian tumors and found that this gene was frequently inactivated. We also found that re-introduction of Parkin is associated with greater sensitivity to apoptotic induction in ovarian cancer cell lines. We also discovered an entire family of very large common fragile site genes. We measured the expression of Parkin and 13 other CFS genes in panels of different cancers. This revealed non-random inactivation of these genes and greater inactivation in cancers that have a poorer clinical prognosis. We then utilized whole genome tiling arrays to characterize all transcripts (not just coding transcripts) and their response to stress. These studies revealed non-coding transcripts within Parkin and other large CFS genes. We discovered a new class of large non-coding transcripts. A sub-set of these were highly evolutionarily conserved; when we examined these transcripts in ovarian and other cancers, we found alterations in their expression. We also found that some of these non-coding transcripts were mutational targets in ovarian and other cancers. Thus, this work has discovered a new group of targets that are altered during the development of ovarian cancer.

Table of Contents

Page

INTRODUCTION

Parkin is a gene that spans an extremely large chromosomal region of 1.36 megabases (Mbs). This large gene spans the distal half within the highly unstable FRA6E CFS (6q26). Our novel hypothesis that received funding from the Department of Defense Ovarian Cancer Research Program was that Parkin and other large CFS genes were part of a stress response system disrupted during ovarian cancer development. There were two major goals to this proposal. The first was to determine if inactivation of expression of Parkin and other large CFS genes could contribute to the development of ovarian cancer. The second was to demonstrate that Parkin and other large CFS genes functioned as part of a stress response system within cells. We will now summarize the work that we completed after our third year of Department of Defense funding. Our key findings were as follows: (1) We demonstrated that Parkin is indeed inactivated in many ovarian and other cancers and that the re-introduction of Parkin results in growth inhibition of ovarian cancer cell lines; (2) We identified an entire family of very large CFS genes that have similar structures to Parkin and that are also highly evolutionarily conserved (in spite of the fact that they reside within the highly unstable CFS regions); (3) We showed that these genes are not randomly inactivated in ovarian or other cancers, and that there is a correlation between inactivation of multiple large CFS genes and cancers that are generally associated with a poorer overall clinical outcome; (4) We identified the retinoic acid receptor-related orphan receptor alpha (RORA) as a large CFS gene whose expression is abrogated in many different cancers including ovarian cancer. This nuclear transcription factor is extremely interesting because in addition to regulating many key cellular functions, it also appears to function as a stress regulated gene; (5) We found that many of the large CFS genes appear to play key roles in cellular responses to stress and most particularly oxidative stress; and (6) We initiated a very novel experiment to characterize the large CFS genes as part of a stress response system within cells utilizing the newly developed whole genome tiling arrays.

BODY

We would again like to thank the Department of Defense for their support of our work. We continue to believe that the characterization of the common fragile sites (CFSs) is important because these large regions of genomic instability that are found in all individuals are chromosomal regions that are uniquely sensitive, especially in developing ovarian cancers. In addition, many contain large genes that are particularly attractive tumor suppressor candidates which participate in ovarian cancer development. The first two genes identified within these unstable chromosomal regions were FHIT and WWOX. These two genes have a very unusual genomic organization as they both span extremely large genomic regions (1,2) greater than 1.0 Mb in size. In spite of this, the final processed transcripts of these genes are relatively small (1.0 Kb for FHIT and 2.0 Kb for WWOX); thus, the majority of these genes (greater than 99.9%) are comprised of intronic sequences. These two genes have been demonstrated to function as tumor suppressors both *in vitro* and *in vivo* (3-5). In addition, inactivation of these genes is associated with a poorer clinical outcome (6-9). Finally, both genes are involved in cellular responses to stress (10-12).

The original goal of this proposal was to characterize another large CFS gene, the 1.36 Mb Parkin gene. This gene spans the third most unstable CFS region, FRA6E (6q26) (13). Our two main goals were: (A) to characterize the Parkin gene in ovarian and other cancers and to determine if the inactivation of this gene was a frequent event in ovarian cancer and to determine if this had any functional significance; and (B) To determine if Parkin and other CFS genes were involved in the cellular responses to stress. We completed Specific Aim #1 in the first year and we demonstrated that this gene was indeed frequently inactivated in ovarian cancers, as well as in prostate and breast cancers, and in hepatocellular carcinoma (13-14). We also showed that re-introduction of Parkin into an ovarian cancer cell line that did not produce any endogenous Parkin results in growth inhibition and also protected cells from mitochondria-independent apoptosis induced by ceramide. This work was published in *Oncogene* (13) and *Genes, Chromosomes, and Cancer* (14). Parkin has been found to also be a stress-responsive gene, similar to FHIT and WWOX, as it was recently shown that there is induction of Parkin expression in the presence of oxidative stress (15).

In this report we summarize our work where we have now identified an entire family of extremely large CFS genes. One CFS gene that we have now identified is RORA, the retinoic acid orphan receptor alpha. This 730 Kb gene spans the center of the FRA15A CFS (15q22.2) and is an extremely interesting nuclear transcription factor involved in the regulation of a number of key cellular processes. In addition, this gene is a cellular stress response gene. We describe our work characterizing the RORA gene where we demonstrate that this gene is

frequently inactivated in multiple cancers including ovarian cancer and that it is involved in cellular stress response. The work on RORA was recently published in *Oncogene* (14).

When we originally wrote this proposal, we wanted to characterize how large genes like FHIT, WWOX, and Parkin could be responding to cellular stress. We could not imagine that a powerful technology like tiling arrays would be developed which in a single experiment would enable us to probe the entire genomes response to stress. We have been beta-testing the new 35 bp genome tiling arrays (which contain tiling oligonucleotides spaced 35 bp apart across the entire non-redundant portion of the human genome); we have devised an experiment to measure both coding and non-coding transcripts across the entire genome and their response to two types of stress, growth under hypoxic conditions, and exposure to the carcinogen NNK. We have completed this experiment and are beginning to analyze the huge amount of data generated. In the next year, we will be able to determine whether these stresses cause changes in non-coding transcripts which are present within the large introns of CFS genes like Parkin. Our hypothesis is that the reason there are such large genes within the highly unstable CFS regions is that the CFS regions are able to somehow transduce different cellular stresses into the production of the appropriate non-coding transcripts which then regulate the expression of the large CFS genes.

In this report, we therefore summarize our work on the identification of an entire family of large CFS genes, the identification and analysis of the RORA gene, and our preliminary studies utilizing tiling arrays to characterize the entire genomes response to stress.

Large genes within many CFS regions

Our analyses of several CFSs revealed that there were large genes (genes >1.0 Mbs) located within several, but not all, of these regions including FHIT (1.5 Mbs), Parkin (1.36 Mbs), GRID2 (1.39 Mbs), and WWOX (1.0 Mbs). Additionally, along with other,s we demonstrated that FHIT, WWOX, and GRID2 were highly evolutionarily conserved and that the chromosomal regions surrounding them were also CFSs in mice (15-17). This suggested that the large gene and the unstable chromosomal region might be co-conserved because together they serve some function within cells.

We became interested in whether other CFS regions might also include large genes. To address this question, we collaborated with Dr. Robert Kuhn, a researcher at the UCSC Genome Database. Dr. Kuhn provided a list of all genes larger than 500Kb, and we carefully examined the list to remove redundant clones. We generated a list of 240 distinct human genes that spanned greater than 500 Kb of genomic sequence. These 240 genes represent the largest 1% of human genes.

Many of the largest human genes are derived from within CFS regions

Examination of the large gene list revealed that a number of these were derived from chromosomal bands that contained CFSs; we were curious how many corresponded to CFS genes. Our laboratory had already localized 20 of the 89 known CFS regions, and a few other CFS regions have been defined by other groups (18-21). A detailed examination of the sequences surrounding these localized CFSs identified several other large genes as CFS genes, including CNTNAP2 [the largest human gene which spans 2.3 Mb within 7q35 (FRA7I)] and LRP1B (1.9 Mb in FRA2F).

We then decided to test several of the largest human genes derived from chromosomal regions known to contain a CFS to determine if they were also CFS genes. A BAC clone covering the approximate center of each large gene was selected, labeled, and used as a FISH probe against metaphases prepared from cells exposed to aphidicolin. This analysis identified several other large CFS genes. However, not every large gene was derived from within a CFS region. Out of the 10 largest genes, 6 were determined to be derived from within CFS regions. Closer examination of the genomic region surrounding many of the localized CFS regions revealed that slightly less than half of the characterized CFS regions are associated with large genes. We can therefore estimate that there are approximately 40 large CFS genes distributed throughout the genome (22).

The following Table lists the 20 known large CFS genes that have been identified as of today, the size of the genomic region spanned by each gene, the number of exons and the size of the final processed transcripts (FPT), the chromosomal location, and the CFS that spans each gene.

Gene Name	Size	Exons/FPT	Location	Fragile Site
CNTNAP2	2304258	25/8107	7q35	FRA7I
DMD	2092287	79/13957	Xp21.1	FRAXC
LRP1B	1900275	91/16556	2q22.1	FRA2F
CTNNA3	1775996	18/3024	10q21.3	FRA10D
DAB1	1548827	21/2683	1p32.3	FRA1B
FHIT	1499181	9/1095	3p17.2	FRA3B
KIAA 1680	1474315	11/5803	4q22.1	FRA4D
GRID2	1467842	16/3024	4q22.3	FRA4D
Dlg2	1463760	23/3071	11q14.1	FRA11F
Parkin	1379130	12/2960	6q26	FRA6G
IL1RAPL1	1368739	11/2722	Xp21.2	FRAXC
WWOX	1113013	9/2264	16q23.2	FRA16D
PDGFFA	917434	24/2550	4q12	FRA4B
IMMPL2	899238	6/1540	7q31.1	FRA7K
RORA	731967	11/1816	15q22.2	FRA15A
PTPR6	731390	30/4707	3p14.2	FRA3B
Neurobeachin	730417	58/10812	13q13.2	FRA13A
LARGE	647480	16/4326	22q12.3	FRA22B
ARHGAP15	638958	14/1706	2q22.2	FRA2F
SCAl	462345	9/10601	6p22.3	FRA6C

Similarities between the known large CFS genes

The large CFS genes share a number of similarities. Each of these genes is predominantly intronic (greater than 99.7%) and span some of the most unstable chromosomal regions in the genome which are difficult regions to transcribe as well as replicate. In addition, several of the large genes such as FHIT, WWOX, and GRID2 have been found to be highly conserved and the chromosomal regions surrounding them are fragile sites in mice. When comparing what little is known about the function of some of these large genes, it appears that many of them have completely different functions. However, one interesting connection shared by many of the large CFS genes is an association with normal neurological development.

This is already quite clear with Parkin, which when inactivated results in specific cellular death of cells that make and respond to dopamine leading to early onset juvenile Parkinson's disease (23). A spontaneous mouse mutant was identified that had a 1.0 Mb deletion within the distal portion of FRA6E. This deletion removed coding sequences of both Parkin and the immediately adjacent large Parkin co-regulated gene product PACRG. The mice that are homozygous for this deletion have a neurological phenotype known as *Quaking*(viable) (24), which have quake-like tremors caused by improper myelination of the CNS.

The GRID2 gene is a very large CFS gene which was first identified because of spontaneous deletions in mice resulting in a neurological defect known as *Lurcher* (25). Heterozygous *Lurcher* mice display ataxia as a result of selective, cell autonomous and apoptotic death of cerebellar Purkinje cells during postnatal development (17). This gene is also highly conserved between humans and mice and the region surrounding this gene is a CFS in the mouse (17), identical to what was observed for FHIT and WWOX.

Yet another large gene we identified as a CFS gene was DAB1. Dab1 is the human homolog of the Drosophila disabled locus and it interacts with Reelin. When this locus is mutated in mice, it results in a neurological defect known as *scrambler* (26). These mice have cerebellar hyperplasia with Purkinje cell ectopia. The normal function of Dab1 is to promote normal positioning of upper layer cortical plate neurons (27).

We tested a number of the largest human genes that were involved in neurological development, which were also derived from chromosomal regions known to contain a CFS and identified a number of other large CFS genes. Included were the Duchene Muscular Dystrophy (DMD) gene and the immediately adjacent large

Il1RAPL1 gene in FRAXC, LARGE in FRA22B (which is associated with myodystrophy when deleted in the mouse), and the Sca1 gene (associated with spinocerebellar ataxia) in FRA6C. Thus, many CFS genes are large genes involved in normal neurological development.

RORA is another large CFS gene

In the second year of this work, we began to characterize another very interesting gene, the 730 Kb retinoic acid related orphan receptor alpha (RORA) which was localized to a chromosomal band 15q22.2 that contained another CFS region, FRA15A. We demonstrated that this gene was localized within the middle of this CFS region. However, unlike our previous studies with Parkin, we did not completely characterize the FRA15A CFS region. Our reason for this was that FRA15A is a CFS region that is not frequently expressed (as are FRA3B, FRA16D and FRA6E); thus we would have had to analyze thousands of metaphase spreads to find sufficient metaphases with breakage/decondensation in the FRA15A region to fully characterize this CFS region. We then analyzed expression of RORA in ovarian cancers as well as in several additional cancers and found that this gene, like Parkin, is frequently inactivated in ovarian (and other) tumors. To determine if this had any functional significance, we over-expressed RORA in a breast epithelial cell line and found that this resulted in growth inhibition of the cells (14).

RORA is a particularly interesting gene because it is a nuclear transcription factor that has already been linked to a number of critical cellular functions including control of circadian rhythm, control of cholesterol metabolism and it binds directly to hypoxia inducible factor 1 alpha (28). In addition, RORA is a stress-responsive gene whose expression increases in response to growth under hypoxic conditions, as well as exposure to aphidicolin and chemotherapeutic agents used in the treatment of ovarian cancer. Another interesting thing about RORA is that like Parkin is associated with a neurological defect. Mice that have spontaneous deletions in RORA have the neurological defect *staggerer*, which results from selective death of Purkinje cells (30). This turns out to be a common theme with a number of the large CFS genes, namely that defects in these genes are associated with neurological development and also with the development of cancer.

Because RORA shares similarities with other well studied large CFS genes and is located in a chromosomal band (15q22.2) known to contain the FRA15A CFS, we tested whether RORA was in fact a CFS gene. This analysis revealed that RORA spanned the middle of the FRA15A region. Figure 1 below shows representative FISH results with a BAC from the middle of the RORA gene demonstrating that in one metaphase (A), the BAC hybridizes distal to the region of decondensation/breakage and in another metaphase (B), the BAC hybridizes proximal.

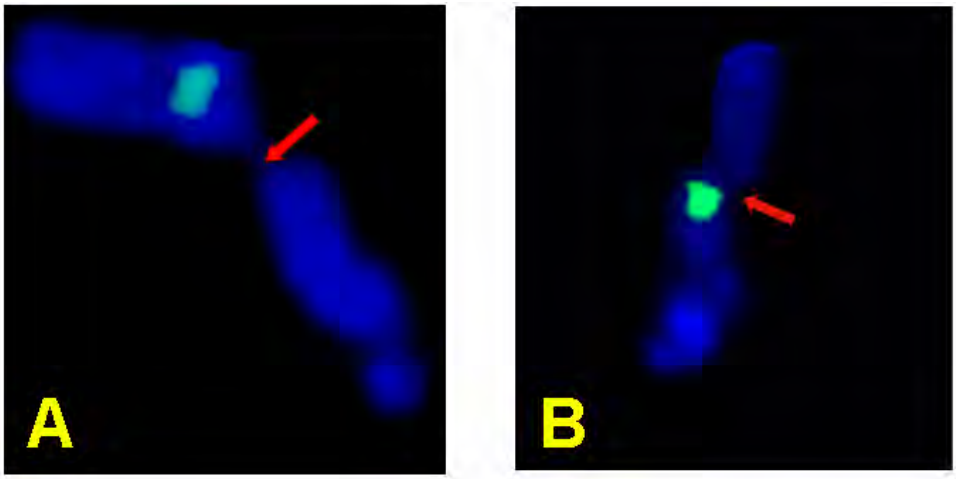


Figure 1 Depiction of FISH results obtained with a BAC clone crossing the middle of RORA and determined to be crossing FRA15A. BAC clone CTD-2034M3 was labeled with biotin and hybridized to normal human lymphocytes treated for 24 hours with 0.4μM aphidicolin. 20 metaphases with clear breakage/decondensation at 15q22.2 were scored. The hybridization signal appeared proximal to the break in 12 metaphases and distal in 8, showing that RORA is located in the approximate center of FRA15A. A. Representative metaphase with the hybridization signal appearing distal to the break. B. Representative metaphase with the hybridization signal appearing proximal to the break.

Out of 20 metaphases with good discernible breakage within 15q22.2, we found that this BAC hybridized proximal to the region of breakage 12 times and distal 8 times. This finding would place this BAC clone and the RORA gene itself within the middle and most unstable region of FRA15A.

According to one previously published study, there are four RORA isoforms (RORA 1, 2, 3, and 4) which are produced by alternative splicing. In our studies in various normal tissues, we found expression of only RORA 1 and 4. Figure 2 below shows the transcriptional level of RORA 1 and 4 in various normal tissues.

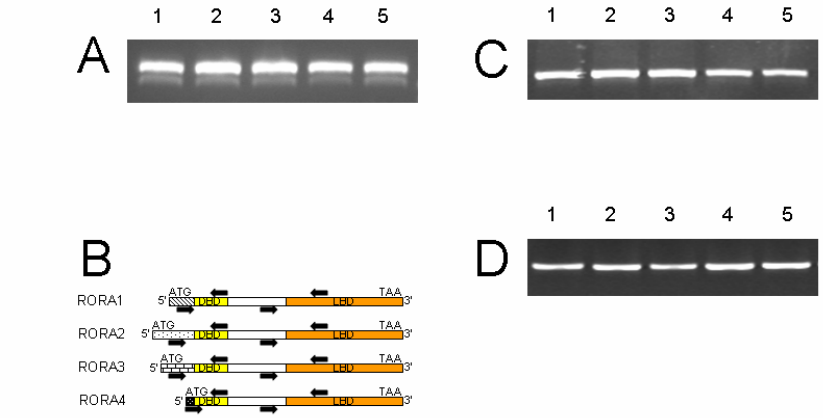


Figure 2 The transcriptional level of RORA and its isoforms in various normal tissues. Lane 1 brain; Lane 2 breast; Lane 3 liver; Lane 4 ovary; Lane 5 prostate. Total RNA was prepared from normal human tissues and cDNA was generated. Semi-quantitative RT-PCR was performed using the universal primers for all RORA isoforms and the specific primers for each isoform to measure the level of RORA. A. RORA universal primers; B. A schematic diagram showing the four different isoforms of RORA. The bold arrows show the positions of specific primers for each isoform and the universal primers for all isoforms. DBD DNA binding domain; LBD ligand binding domain. C. RORA1 (isoform 1) primers; D. RORA4 (isoform 4) primers.

We next examined the expression of RORA in several different types of human cancer samples, either in primary tumors or in tumor-derived cell lines using RT-PCR. This revealed that RORA was down-regulated in breast, prostate, and **ovarian** cancers (see Figure 3 below). These results are consistent with those obtained in studies of other critical CFS genes including FHIT, WWOX, and Parkin.

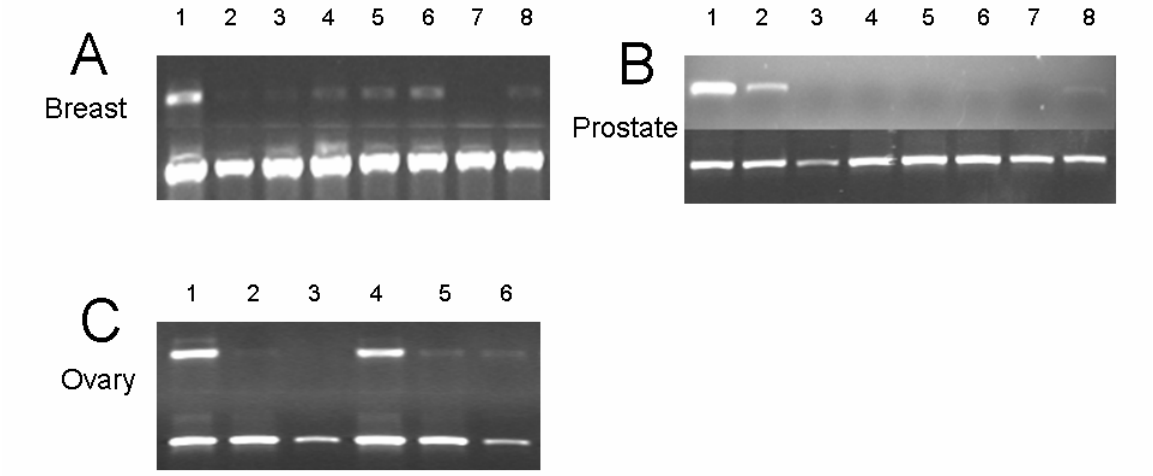


Figure 3. RORA is down-regulated in different types of cancer cell lines and human primary cancers. Total RNA was extracted and reverse transcribed into cDNA. PCR was performed with RORA universal primers to examine the transcriptional level of RORA. A. Breast cancer cell lines, Lane 1 MCF12F; Lane 2 MCF7; Lane 3 MDA157; Lane 4 UACC893; Lane 5 ZR75; Lane 6 MDA435; Lane 7 T47D; Lane 8 BT474. Top row RORA; Bottom row Actin. B. Prostate cancer cell lines and primary tumor samples, Lane 1 normal prostate control; Lane 2 DU145; Lane 3 PC3; Lane 4 LNCaP; Lane 5-8 primary prostate tumor tissues. Top row RORA; Bottom row Actin. C. Ovary cancer cell lines, Lane 1 normal ovarian epithelium control (OSE); Lane 2 OV167; Lane 3 OV177; Lane 4 OV202; Lane 5 OVCAR5; Lane 6 SKOV3. Top row RORA; Bottom row Actin.

We also examined whether RORA expression was modulated by different types of cellular stress other than hypoxia. We first demonstrated that RORA is activated by exposure to aphidicolin (Figure 4) and then subsequently showed similar activation by other types of cellular stress including exposure to ultraviolet radiation (UV), the addition of the carcinogen MMS (methyl-methane sulfonate), and the treatment with H_2O_2 (oxidative stress) (2). Figures 4 and 5 display the changes in RORA transcripts and RORA protein levels in response to some of these stresses.

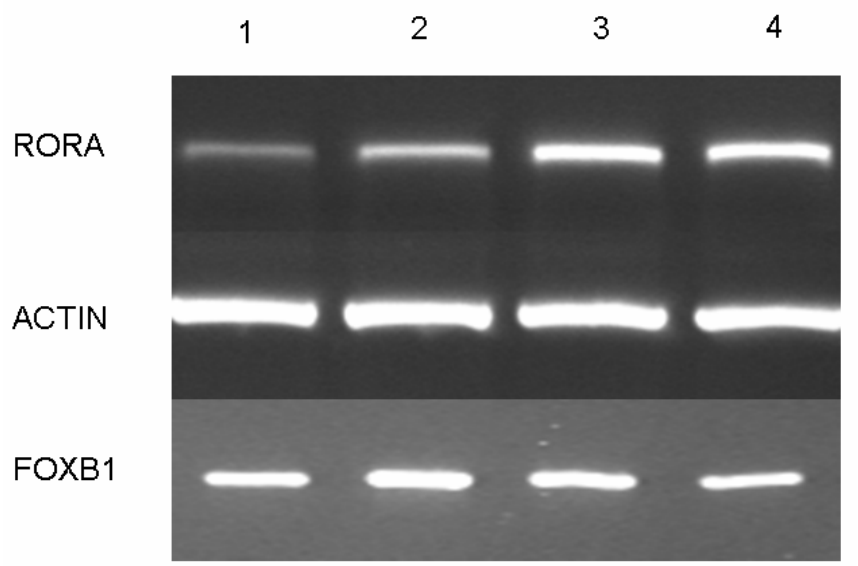


Figure 4. The effect of aphidicolin (APC) on the transcription of RORA in MCF12F cells. MCF12F cells were treated with various doses of APC for 24 hours before total RNA was extracted and cDNA was prepared. PCR was set up to check the transcriptional level of RORA, FOXB1 (a gene within FRA15A right next to RORA) and Actin, using the universal primers for RORA, primers for FOXB1 and the control primers for Actin. Lane 1 cell without APC treatment; Lane 2 with APC 0.2 μM; Lane 3 with APC 0.4 μM; Lane 4 with APC 0.8 μM.

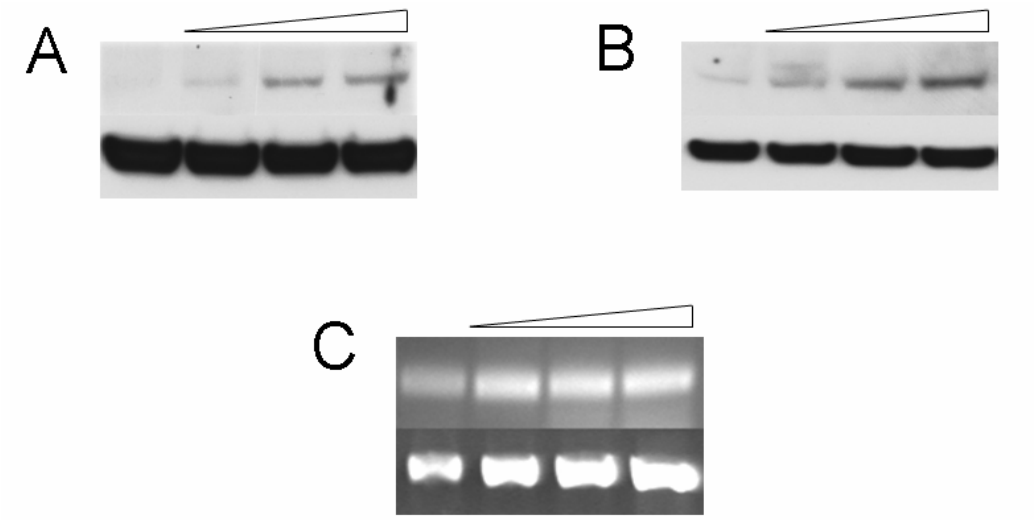


Figure 5. The expression of RORA in MCF12F cells is activated by different types of stress treatments. A. The effect of UV on the protein level of RORA. MCF12F cells were treated with UV at 10, 20 and 50 J/m^2 and total protein was prepared. B. The effect of MMS on the protein level of RORA. MCF12F cells were treated with MMS at 0.001%, 0.005% and 0.01% for 24 hours. The level of Rora was examined with anti-Rora antibody. C. The effect of H_2O_2 on the transcriptional level of RORA. MCF12F cells were treated with H_2O_2 at 100, 200 and 500 μM for 24 hours and then total RNA was extracted and cDNA was prepared. RT-PCR was performed using universal primers for RORA.

An important question is what role alterations in expression of this large CFS gene play in the development of cancer? Indeed all of the CFS genes could be frequent targets of alterations in unstable cancer cells because of the unstable regions that surround them. We transfected RORA into the breast cell line MCF12F and found that increased RORA expression resulted in decreased growth of MCF12F cells (see Figure 6 below). These results are similar to those obtained with FHIT, WWOX, and Parkin.

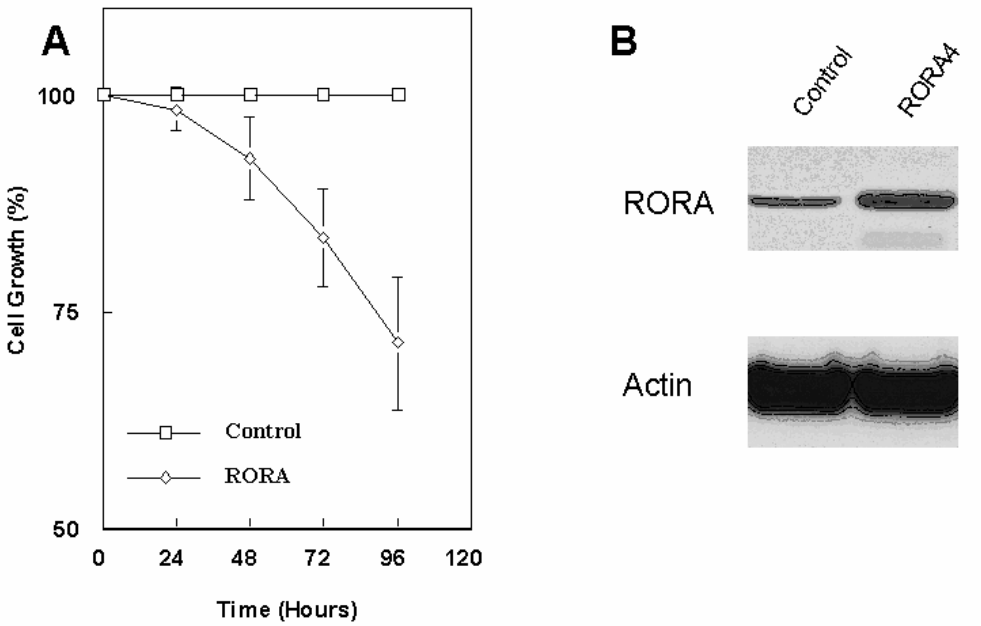


Figure 6. The effect of RORA over-expression on cellular growth. A. MCF12F cells were plated and incubated overnight before transfection. The plasmids (pcDNA3 as control and pcDNA3-RORA4) were transfected into cells using a Lipofectamine 2000 transfection kit following the manufacturer's protocol. The cell number was counted 24, 48, 72 and 96 hours later. All results are the average of at least three independent experiments with standard deviations shown by bars. B. The level of RORA in pcDNA3-RORA4 transfectants detected by the Western blotting assay. Top row RORA; Bottom row Actin.

Thus, changes in RORA expression are associated with readily observable changes in growth rates of MCF12F cells, which supports our contention that inactivation of RORA expression could provide a significant growth advantage to cells thus participating in breast cancer development. All of this work is summarized in our recently published *Oncogene* paper (29).

Expression of the large CFS genes in cancers and cancer-derived cell lines

Our hypothesis is that the large CFS genes are part of a stress response system within cells that is uniquely susceptible to genomic instability and that in cancers with considerable genomic instability, there will be inactivation (alterations) of expression of multiple CFS genes. We have already demonstrated observable phenotypic changes associated with alterations in the expression of these genes. Next we sought to determine whether these genes were randomly inactivated or whether there might be some selection for inactivation of specific CFS genes in different cancers.

To address this question, we used real-time RT-PCR to precisely measure the expression of seven representative large CFS genes (FHIT, WWOX, **Parkin**, GRID2, DLG2, DAB1 and the two expressed RORA isoforms 1 and 4) in panels of primary tumors and cancer-derived cell lines for cancers of the prostate, breast, ovary, liver, and brain. PCR primers were constructed to be optimal for real-time RT-PCR analysis (100-125 bp products derived from the 3' end of the final processed transcripts from these genes) and then we performed real-time RT-PCR in the ABI 7900 real-time PCR machine. To quantify the expression of each of these genes (we constructed primers to differentiate between the two RORA isoforms), we compared the C_t measurements obtained with each gene to that of the β–actin gene and used the delta C_t measurements to quantify message amounts for the large CFS genes. We obtained several normal tissues for comparison for each tissue/tumor type and compared the expression of β-actin and the CFS genes in those normal tissues to panels of cancer-derived cell line, as well as primary tumors of that same type. We considered any gene to be aberrantly regulated if its expression was more than 4-fold up or down relative to the range of expression determined for the normal samples after each sample was run in triplicate.

We found that the expression of these genes was frequently abrogated in different cancers and there appeared to be a very non-random pattern of gene inactivation. We also observed that many cancers had inactivation of multiple large CFS genes. Those cancers with a great deal of genomic instability will have inactivation of many of these genes simultaneously which could have a profound phenotypic effect on those cells. For each of the

CFS genes tested, the Table below indicates the number of primary tumors/cell lines that had decreased expression compared to normal samples divided by the total number of tumors/cell lines tested.

	FHIT	WWOX	**Parkin**	Grid2	Dlg2	Dab1	RORA1	RORA4
Prostate	0/17	1/17	1/17	0/17	2/17	0/17	10/17	1/17
Ovary	**2/18**	**1/18**	**3/18**	**1/18**	**12/18**	**0/18**	**2/18**	**5/18**
Breast	4/16	3/16	5/16	0/16	8/16	3/16	3/16	**8/16**
Brain	7/17	10/17	5/17	9/17	17/18	11/17	5/17	0/17
Liver	4/15	11/15	14/15	14/15	12/15	12/15	15/15	0/15

It is important to note that there was a very interesting preliminary correlation between the frequency of inactivation of these CFS genes and cancers that have very poor clinical outcomes. We found the least inactivation of CFS gene expression in cancers of the prostate, which of the various cancers examined has the best clinical outcome. There was much greater loss of expression of these genes in cancers of the breast and ovary, and many of these cancers tend to be more aggressive than prostate cancers. However, the cancers with the greatest inactivation of these genes were cancers of the brain and liver. These cancers are highly aggressive, and there is a high probability that patients who develop these tumors will succumb to them.

We have subsequently done similar measurements but this time examined a total of 14 large CFS genes. The results obtained were very similar to what was observed with the smaller group of genes. We now have several publications examining the expression of this panel of large CFS genes in different cancers (see *Publications as a Direct Result of this Grant Support* at the end of the report). What we have demonstrated is that there is non-random inactivation of these genes in different cancers. In addition, genes that reside within the most unstable CFS regions are not necessarily the genes that show the greatest frequency of loss of expression in different cancers. This argues that there is some sort of selection for loss of expression of different CFS genes in different cancers.

Disabled-1 (Dab1) is another large CFS gene

Another large gene that we began to examine was the human homolog of the disabled locus from Drosophila, Dab-1. This gene spans 1.25 Mbs within chromosomal band 1p32.2, which is also the band that contains the FRA1B CFS. We were, therefore, interested in whether Dab1 was another large CFS gene that could be important in cancer development. We found that the large Dab1 gene was derived from within the middle of the FRA1B CFS region. We then used real-time RT-PCR analysis to find that Dab1 was inactivated in several different cancers, including ovarian cancer. However, it was more frequently down-regulated in brain and endometrial cancers. We then found that over-expression of Dab1 resulted in decreased cell growth in two different cell lines. This is similar to what we observed in RORA. Interestingly, when this gene is inactivated in mice, it results in the neurological mutant *scrambler* (26) which has abnormal cortical lamination indistinguishable from another mouse neurological mutant *Reeler*, caused by mutations in the reelin gene. This provides further proof that there is a connection between these large genes involved in neurological development and genes that could play an important role in cancer development.

Monitoring the entire genomes response to stress using tiling arrays

Much of the focus in cancer genetics has been on the identification of important alterations during cancer development which could contribute to that process. This focus has been primarily upon the genes, and more importantly the coding portions of those genes. However, the coding portions of the genome comprises less than 2% of the entire genome. If you subtract the 45% of the genome that is comprised of repetitive DNA that still leaves 53% that is neither coding nor repetitive. Nowhere is this discrepancy between genome size and apparent coding potential more evident than in some of the large CFS genes. The 1.36 Mb Parkin gene produces a 2960 bp final processed transcript, making this gene over 99.8% intronic. Similarly, the 730 Kb RORA gene has a 1095 bp final processed transcript, making this gene also over 99.8% intronic. Why produce such a large initial transcript only to process it down to such a small final processed transcript? One possibility is that are regulatory non-coding RNAs encoded within the intronic sequences which regulate the expression of such large genes like Parkin. However, with no knowledge of where within such large genes these transcripts are derived, there used to be no feasible way to identify these stress-regulated transcripts.

We were considering the construction of sufficient oligonucleotides to probe across the large Parkin introns in order to identify putative non-coding transcripts at a cost of many thousands of dollars when Affymetrix began to beta-test their new whole-genome tiling arrays. The Human Tiling 1.0R Array set contains tiling oligonucleotides spaced 35 bp apart (from the center of one oligonucleotide to the center of the next adjacent oligonucleotide across the entire non-redundant portion of the genome). We obtained multiple sets of these chips as part of the beta-testing of these new arrays from Affymetrix and decided to set up an experiment that would enable us to not only identify transcripts (and non-coding transcripts) across the entire genome but also to ask which of those transcripts were stress-responsive (especially since we believe that the large CFS genes are actually part of a stress-response system within cells).

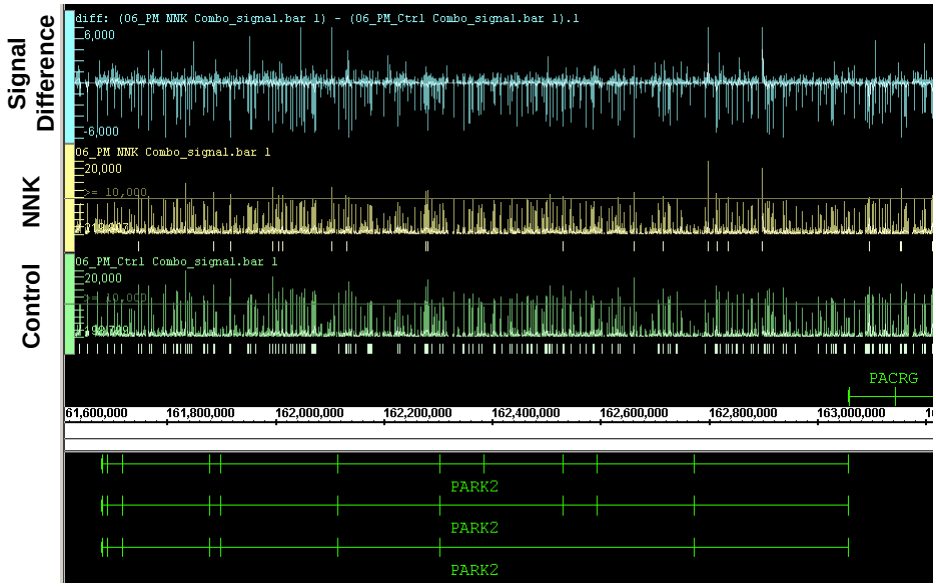
The work with tiling arrays has been pioneered by Dr. Tom Gingeras and co-workers at Affymetrix. Using 5 bp tiling arrays (which have a much higher density of tiled oligonucleotides that are 5 bp apart from center to center of adjacent oligonucleotides), they demonstrated that unannotated, non-polyadenylated transcripts comprise the majority of the transcriptional output for the human genome (29). This provides additional support for our hypothesis that the large introns in the large CFS genes might produce important stress-regulated transcripts which are involved in the stress-response that the large CFS genes are involved.

Our microarray experiment was set up to measure the entire genomes response to two different types of cellular stress: growth under hypoxic conditions and exposure to the carcinogen from cigarette smoke [4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK)] (24). We cultured normal ovarian surface epithelial cells and exposed them to hypoxia or NNK. Hypoxia is a physiologically important endoplasmic reticulum (ER) stress that is present in all solid tumors. In addition, both Parkin and RORA respond to oxidative stress. We chose NNK because it induces DNA damage, feeling that the comparison between the two different types of stress would yield different types of stress-responsive transcripts.

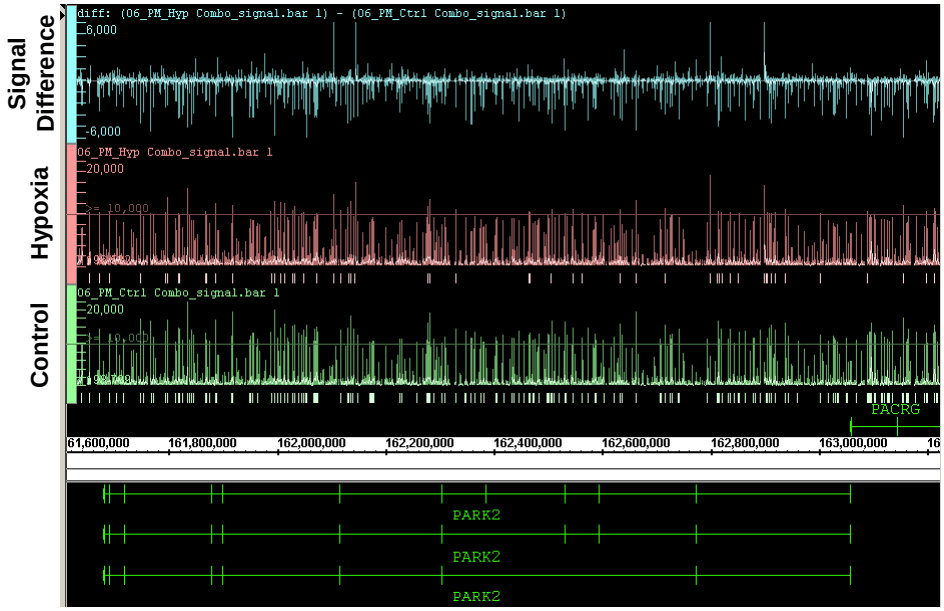
This work was done in collaboration with the Microarray Core of the Mayo Clinic and together, we went through the protocol for producing cDNA from all the RNA species present (not just poly A^+ RNA) using random oligonucleotide primers. One potential problem with this experiment is that total RNA contains a vast excess of ribosomal RNA which could potentially swamp out signals from both coding genes and non-coding transcripts. There is a commercially-available kit using magnetic beads which can purify away the ribosomal RNA (the RiboMinus Kit), and we compared the hybridization of cDNA produced from total RNA to cDNA produced after the use of the RiboMinus kit. The RiboMinus protocol resulted in significant enrichment of non-ribosomal RNA but also resulted in some degradation of the remaining RNA species. We hybridized equivalent amounts of labeled total RNA and RiboMinus purified RNA (-rRNA) to the tiling array chips so that we could compare the hybridization signals to help determine whether the additional expense of the RiboMinus purification kits was worthwhile. We found almost identical hybridization with both sets of RNAs (after being converted to cDNA), and thus did the entire experiment without the RiboMinus purification. We hybridized the three sets of labeled probes (control ovarian surface epithelial cells, cells exposed to hypoxia, or cells exposed to NNK) to two sets of the whole genome 35 bp tiling array each.

In a previous report to the Department of Defense, we described some of our preliminary work where we compared total RNA to –rRNA, as well as our results when we examined the hybridization to a single chip (one containing chromosome 7 and the largest known human gene CNTNAP2). In the past year we have been analyzing the results with the whole genome hybridization (to a total of 14 chips each run in duplicate). We performed two types of analyses as described here. The first was to focus on some of the largest human genes and for this we focused on how oligonucleotides tiled across Parkin and RORA hybridized to the three sources of RNA; we could examine both transcripts that hybridized as well as transcripts that were stress responsive.

The Figure on the following page shows the results obtained when we examined the Parkin gene. The bottom figure (in each case) was the control hybridization. The middle figure is the hybridization to one of the two stresses utilized, and the top figure is the subtraction of transcripts seen after stress by the control transcripts. The top of the figure is the signals obtained comparing NNK exposure to the control. The bottom of the figure shows the different Parkin transcripts that are produced (hatched lines represent exons).

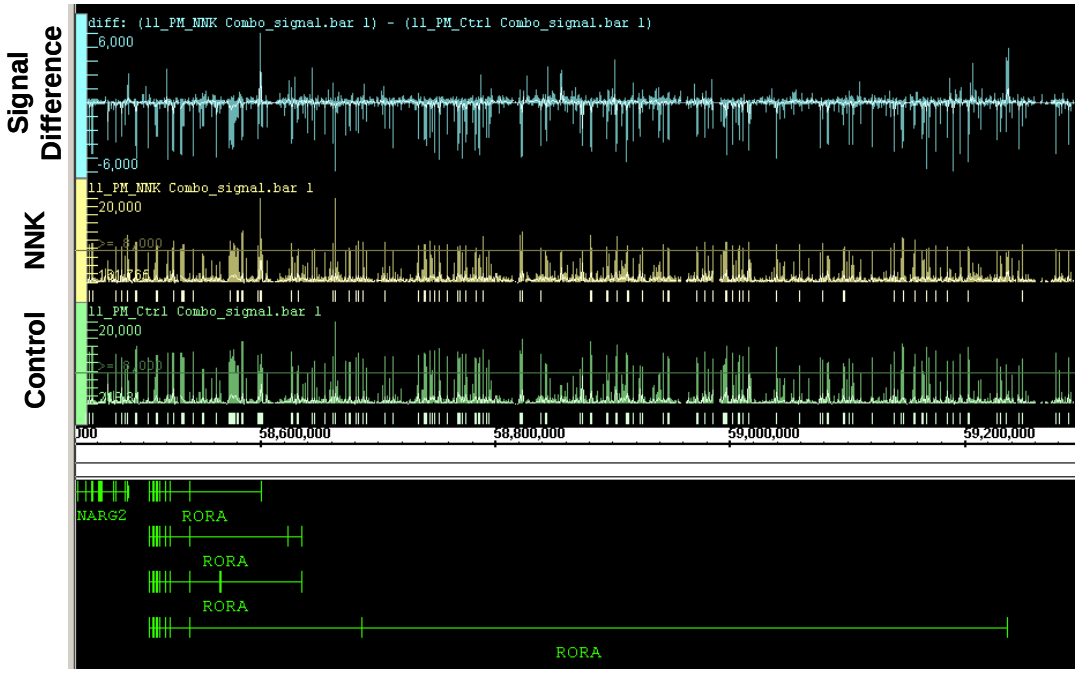


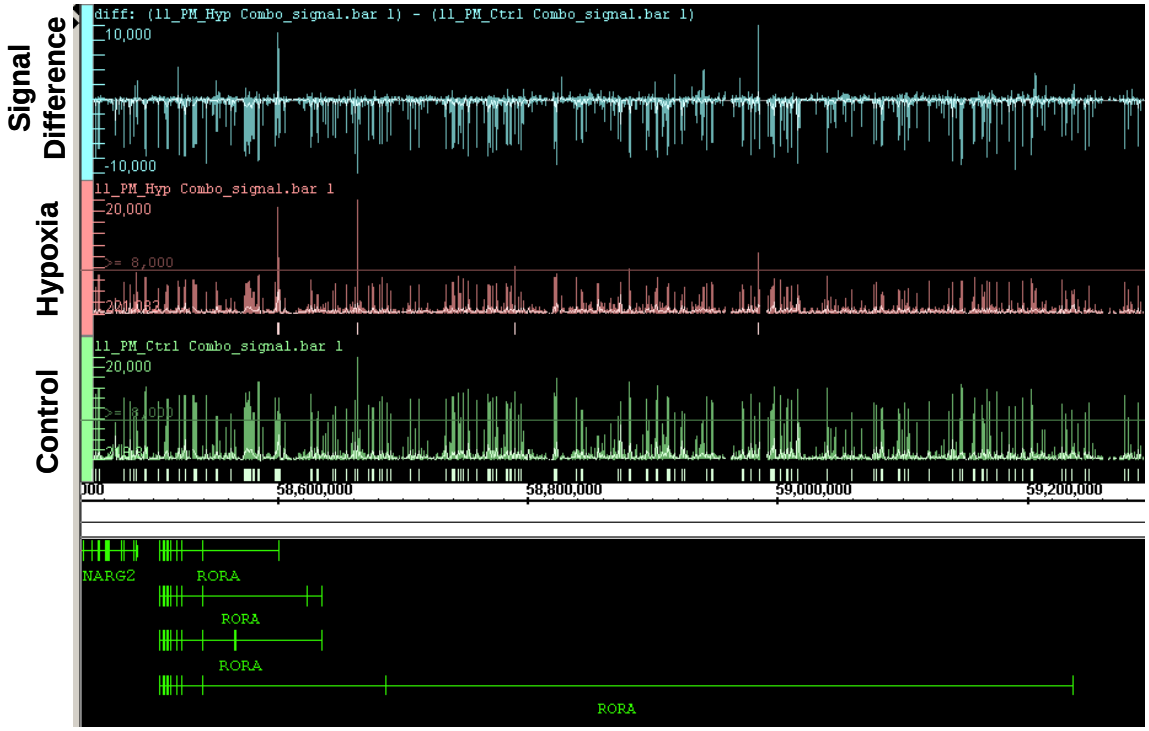
The Figure below shows Parkin comparing the control to hypoxia-induced stress.



Any of the signals that increase in response to the stress are shown as signals above the graph and signals that decrease are shown as signals below the graph. What is very clear from this experiment is that there are multiple transcripts not derived from the exon sequences and that there were many more transcripts which decreased in response to one or both of the stresses applied.

The graphs below show the results that we obtained when we examined the RORA gene.





Large highly conserved non-coding transcripts

The examination of the tiling array data revealed that there were hundreds of thousands of non-coding transcripts across the human genome. In order to focus on a much smaller group of transcripts to further characterize, we first examined only those transcripts that were abundantly expressed. Taking an arbitrary cut-off of 99.5%, we examined transcripts that were highly expressed in the NHBE cells. Transcripts with this high level of transcription had similar levels of expression as actin. We found over 500 transcripts across the genome which were this highly expressed. We then examined each of these 500 transcripts to determine if they were capable of coding for protein and for those that were not, where they were derived from (intronic versus intra-genic). We also examined each potential non-coding transcript to see how evolutionarily conserved they were. Through this analysis we were able to identify a group of large (greater than 400 bp) transcripts that were also highly evolutionarily conserved.

Much of the work over the past year has been on the characterization of these abundantly expressed, highly conserved, large non-coding transcripts. We examined their expression in different normal human tissues and also tested their expression in various cancer derived cell lines and primary tumors. We found that many of them had aberrant expression in both ovarian and breast cancers. In addition, some of them were mutational targets in both of these cancers. We recently published a paper describing all these interesting findings in the journal *Human Molecular Genetics* (30).

KEY RESEARCH ACCOMPLISHMENTS

We have made several major research accomplishments during the course of this Department of Defense grant. The most important is the identification of an entire family of extremely large CFS genes. There are now 20 known large CFS genes, and we estimate that there may be 45-50 all together. We have found and studied a very interesting stress responsive gene, RORA. This important nuclear transcription factor is involved in the regulation of a number of key cellular processes, and we have shown that the expression of this gene is frequently inactivated in many ovarian cancers. Interestingly, this gene, which is similar to Parkin, is involved in cellular responses to oxidative stress. We have also shown that the expression of the large CFS genes is frequently inactivated in multiple tumor types and that there is more inactivation of their expression in cancers associated with a poorer overall clinical outcome (such as brain or liver cancer). We have identified a family of new non-coding transcripts which reside within the two large stress-responsive CFS genes we are currently studying. Our next studies will be to characterize these non-coding transcripts to determine what role, if any, they play in the development of ovarian cancer and how they are involved in the oxidative stress response pathways that Parkin and RORA participate in. Over the final year of this grant (a no-cost extension), we have also characterized a new group of very large highly conserved non-coding transcripts which are targets of alteration (both in terms of expression and as mutational targets) in ovarian and other cancers. This is extremely important because it helped to identify a whole new group of alteration targets during ovarian cancer development.

PUBLICATIONS AS A DIRECT RESULT OF THIS GRANT

1) Denison SR, Wqang F, Becker NA, Schule B, Kock N, Phillips LA, Klein C, **Smith DI**. Alterations in the common fragile site gene Parkin in ovarian and other cancers. *Oncogene* 2003, 22: 8370-8378.
2) Wang F, Denison S, Lai JP, Phillips LA, Montoya D, Kock N, Schule B, Klein C, Shridhar V, Roberts LR, **Smith DI**. Parkin gene alterations in hepatocellular carcinoma. *Genes, Chromosomes Cancer* 2004, 40: 85-96.
3) **Smith DI**, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett* 2006; 232: 48-57.
4) Zhu Y, McAvoy S, Kuhn R, **Smith DI**. RORA, a large common fragile site gene, is involved in cellular stress response. *Oncogene* 2006; 25: 2901-2908.
5) Reshmi SC, Huang X, Schoppy DW, Black RC, Saunders WS, **Smith DI**, Gollin SM. Relationship between FRA11F and 11q13 amplification in oral cancer. *Genes Chromosomes Cancer* 2007; 46: 143-154.
6) **Smith DI**, McAvoy S, Zhu Y, Perez DS. Large common fragile site genes and cancer. *Seminars Cancer Biol* 2007; 17: 31-41.
7) McAvoy, S, Ganapathiraju SC, Ducharme-Smith AL, Pritchett JR, Kosari F, Perez DS, Zhu Y, James CD, **Smith DI.** Non-random inactivation of large common fragile site genes in different cancers. *Cytogenetic Genome Res* 2007; 118: 260-269.
8) McAvoy S, Zhu Y, Perez DS, James CD, **Smith DI**. Disabled-1 is a large common fragile site gene, inactivated in multiple cancers. *Genes Chromosomes Cancer*2008; 47: 165-174.
9) Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, Streng PS, **Smith DI**. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet* 2007; Epub ahead of print.
10) McAvoy S, Ganapathiraju S, Perez DS, James CD, **Smith DI**. DMD and IL1RAPL1: Two large adjacent genes localized within a common fragile site (FRAXC) have reduced expression in cultured brain tumors. *Cytogenet Genome Res* 2008, In Press.

REFERENCES

1) Zimonjic DB, Druck T, Ohta M, Kastury K, Croce CM, Popescu NC, Huebner K. Positions of chromosome 3p14.2 fragile sites (FRA3B) within the FHIT gene. *Cancer Res*1997; 57: 1166-1170.
2) Bednarek AK, Laflin KJ, Daniel RL, Liao Q, Hawkins KA, Aldaz CM. WWOX, a novel WW-domain containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res* 2000, 60: 2140-2145.
3) Siprashvili Z, Sozzi G, Barnes LD, McCue P, Robinson AK, Eryomin V, Sard L, TAgliabue E, Greco A, Fusetti L, Schwartz G, Pierotti MA, Croce CM, Huebner K. Replacement of FHIT in cancer cells suppresses tumorigenicity. *Proc Natl Acad Sci* 1997; 94: 13771-13776.
4) Dumon KR, Ishii H, Fong LY, Zanesi N, Fidanza V, Mancini R, Vecchione A, Baffa R, Trapasso F, During MJ, Huebner K, Croce CM. FHIT gene therapy prevents tumor development in Fhit-deficient mice. *Proc Natl Acad Sci USA* 2001; 98: 3346-3351.
5) Fabbri M, Iliopoulos D, Trapasso F, Aquelin RI, Cimmino A, Zanesi N, Yendamuri S, Han SY, Amadori D, Huebner K, Croce CM. WWOX gene restoration prevents lung cancer growth in vitro and in vivo. *Proc Natl Acad Sci USA* 2005; 102: 15611-15616.
6) Lee JL, Soria JC, Hassan K, Liu D, Tang X, El-Naggar A, Hong WK, Mao L. Loss of Fhit expression is a predictor of poor outcome in tongue cancer. *Cancer Res* 2001: 61: 837-841.
7) Toledo G, Sola JJ, Lozano MD, Soria E, Pardo J. Loss of FHIT expression is related to high proliferation, low apoptosis and worse prognosis in non-small-cell lung cancer. *Mod Pathol* 2004; 17:440-448.
8) Guerin LA, Hoffman HT, Zimmerman MD, Robinson RA. Decreased fragile histidine triad gene expression is associated with worse prognosis in oral squamous carcinoma. *Arch Pathol Lab Med* 2006; 130: 158-164.
9) Nunez MI, Rosen DG, Ludes-Meyers JH, Abba MC, Kil H, Page R, Klein-Szanto AJ, Godwin AK, Liu J, Mills GB, Aldaz CM. WWOX protein expression varies among ovarian carcinoma histotypes and correlates with less favorable outcome. *BMC Cancer* 2005; 27: 64.
10) Ishii H, Wang Y, Huebner K. A Fhit-ing role in the DNA damage checkpoint response. *Cell Cycle* 2007; 6: 1044-1048.
11) Ishii H, Mimori K, Inageta T, Murakumo Y, Vecchione A, Mori M, Furaukawa Y. Components of DNA damage checkpoint pathway regulate UV exposure-dependent alterations of gene expression of FHIT and WWOX at chromosome fragile sites. *Mol Cancer Res* 2005; 3: 130-138.
12) Chang NS, Doherty J, Ensign A, Lewis J, Heath J, Schultz L, Chen ST, Oppermann U. Molecular mechanisms underlying WOX1 activation during apoptotic and stress responses. *Biochem Pharmacol* 2003; 66: 1347-1354.
13) Denison SR, Callahan G, Becker NA, Phillips LA, Smith DI. Alterations in the common fragile site gene Parkin in ovarian and other cancers. *Oncogene* 2003; 22: 8370-8378.
14) Zhu Y, McAvoy S, Kuhn R, Smith DI. RORA, a large common fragile site gene, is involved in cellular stress response. *Oncogene*2006; 25: 2901-2908.
15) Sirashi T, Druck T, Mimori K, Flomenberg J, Berk L, Alder H, Miller W, Huebner K, Croce CM. Sequence conservation at human and mouse orthologous common fragile site regions, FRA3B/FHIT and Fra14A2/Fhit. *Proc Natl Acad Sci USA* 2001; 98: 5722-5727.
16) Krummel KA, Denison SR, Calhoun E, Phillips LA, Smith DI. The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. *Genes Chromosomes Cancer*2002; 34: 154-167.
17) Rozier L, El-Achkar E, Apiou F, Debatisse M. Characterization of a conserved aphidicolin-sensitive common fragile stie at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene* 2004; 16: 32-40.
18) Mishmar D, Rahat A, Scherer SW, Nyakatura G, Hinzmann B, Kohwi Y, Mandel-Gutfroind Y, Lee JR, Drscher B, Sas DE, Margalit H, Platzer M, Weiss M, Tsui LC, Rosenthal A, Kerem B. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc Natl Acad Sci USA* 1998; 95: 8141-8146.
19) Morelli C, Karayianni C, Mungall AJ, Thorland E, Negrini M, Smith DI, Barbanti-Brodano G. Cloning and characterization of the common fragile site FRA6F harboring a replicative senescence gene and frequently deleted in human tumors. *Oncogene* 2002; 21: 7266-7276.
20) Curatolo A, Limongi ZM, Pellicia F, Rocchi A. Molecular characterization of the human common fragile site FRA1H. *Genes Chromosomes Cancer*2007; 46: 487-493.

21) Savelyeva L, Sagulenko E, Schmitt JG, Schwab M. The neurobeachin gene spans the common fragile site FRA13A. *Hum Genet* 2005; 19: 2715-2726.
22) Smith DI, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett* 2006; 232: 48-57.
23) Kitada T, Asakawa S, Hattori N Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimuzu N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998; 392: 605-608.
24) Lockhart PJ, O-Farrell CA, Farrer MJ. It's a double knock-out! The quaking mouse is spontaneous deletion of parkin and parkin co-regulated gene (PACRG). *Mov Disord* 2004; 19: 101-104.
25) Lalouette A, Guenet JL, Vriz S. Hotfoot mouse mutations affect the delta 2 glutamate receptor gene and are allelic to lurcher. *Genomics*1998; 50: 9-13.
26) Sheldon M, Rice DS, D-Archelango G, Yoneshima H, Nakajima K, Mikoshiba K, Howell BW, Cooper JA, Goldowitz D, Curran T. Scrambler and yotari disrupt the disabled gene and produce a reeler-like phenotype in mice. *Nature* 1997; 389: 730-733.
27) Rice DS, Sheldon M, D'Archelango G, Nakajumi K, Goldowitz, Curran T. Disabled-1 acts downstream of Reelin in a signaling pathway that controls laminar organization in the mammalian brain. *Development* 1998; 125: 3719-3729.
28) Chauvet C, Bois-Joyeux B, Berra E, Pouyssegur J, Danan JL. The gene encoding human retinoic-acid receptor-related orphan receptor alpha is a target for hypoxia inducible factor 1. *Biochem J* 2004; 384: 79-85.
29) Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 2005; 308: 1149-1154.
30) Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, Ganapathiraju SC, Streng PS, Smith DI. Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Molec Genet* 2007, Epub ahead of print.